Journal of English for Academic Purposes 25 (2017) 71–83

Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

A phraseological exploration of recent mathematics research articles through key phrase frames

Kelly J. Cunningham

Department of English, 206 Ross Hall, 527 Farmhouse Ln., Iowa State University, Ames, IA, 50011-1201, USA

ARTICLE INFO

Article history:
Received 2 May 2016
Received in revised form 19 August 2016
Accepted 17 November 2016
Available online 6 December 2016

Keywords:
Multiword expressions
English for specific purposes
Phrase frames
Mathematical discourse
Academic writing
Mathematics

ABSTRACT

While a wealth of resources is available for teaching research writing of traditional IMRD research papers, instructors have little to draw on when working with graduate students in mathematics. The present study offers insight into recent mathematical research articles through an exploration guided by phrase-frames, recurring multiword units with a variable slot or blank. The paper introduces the concept of key phrase-frames. In doing so, the potential of phrase-frames for understanding subregisters and for teaching is shown. The starting point of this exploration is a 2,289,670-word corpus of 128 recent mathematics research articles collected from eight scholarly mathematics journals. Five- and six-gram phrase-frames were generated in KfNgram. Key phrase frames are identified as those with an occurrence of at least 20 per million words, occurring across 75% of journals and with a sMAPE of 1.95 cut off or higher when compared to the academic section of the Corpus of Contemporary American English as a reference. The 180 resulting frames were coded for open slot position and common pattern before being consolidated and functionally grouped. Core functions related to the aboutness, coherence and moves of the text. The discussion centers on how these groups reveal different aspects of mathematical texts.

1. Introduction

Graduate English for Academic Purposes instructors often work with students from multiple disciplines at once. Previous research has sought to identify and define differences between some of these fields (e.g. Peacock, 2002) by considering for example how knowledge is constructed (e.g. Gray, 2013) or the common field specific multiword expressions (e.g. Hyland, 2008). These studies, however, often fail to provide instructors with adequate guidance for working with students from mathematics.

Instructors who have mathematicians in their graduate writing courses can likely attest that even within the highly controlled genre of published research articles, mathematics seems different. While fine distinctions can be drawn between other scientific disciplines, the publications of many major fields maintain at least the residue of the IMRD structure. Instructors can draw on this structure and the multitude of articles detailing its purposes, moves, and steps to teach a large percentage of their other students. However, when it comes to working with their mathematics students, these instructors are not able to draw from the same depth of research. One of the aims of the present study is to offer these instructors, and those teaching ESP courses for mathematicians, a greater understanding of these mathematics research texts and the language within them from a phraseological standpoint. An additional aim of this study is to introduce and show the promise of a

E-mail address: Kellyc@iastate.edu.

http://dx.doi.org/10.1016/j.jeap.2016.11.005
1475-1585/

phraseological form, the key phrase frame, for research that can aid not only in informing our teaching and understanding of discipline-specific language and texts but also in sparking new areas of research techniques.

1.1. Formulaicity and writing in mathematics

The writing of math is inextricably linked to the doing of math. Inherently contained within a mathematical text, such as a research article, is the math itself and the statements and proofs that represent it. In such, analysis of the language of a mathematics research article is likely to bring to light elements of article writing and also proof and statement writing. It may also be that elements of proof and statement writing carry over to the other areas of the article.

At the time of writing, the language of mathematics research articles has only begun to be linguistically described through a few recent studies. With the notable exception of the work of Ganesalingam (2013) on the syntax and parsing of mathematics and that situated in systemic functional linguistics (Burton & Morgan, 2000; Morgan, 1998), the majority of research on mathematical texts have been limited to applying frameworks from the research articles of other disciplines to those in mathematics (e.g. McGrath & Kuteeva, 2012). The problem arises that frameworks developed for other disciplines often do not transfer well to the domain of mathematics writing. This is easily seen when porting move-step schemes (Swales & Najjar, 1987; Swales, 1990) created for traditional IMRD structured articles. Behnam and Golpour (2014) note that mathematics abstracts avoid the expected use of moves of other research articles. Graves, Moghaddasi, and Hashim (2014) in studying mathematics research article introductions note mathematics writing's deviation from the expected moves and developed a new move—'establishing presumptions'— in response. With the analysis of mathematics research articles largely limited to investigations of abstracts, introductions and stance and limited to borrowing frameworks and categories not designed for the genre, our understanding of these texts and the language that builds them remains incomplete.

Mathematics as a language is often seen to display a preference for a concise symbolic orthography over prose. Additionally, while it has been suggested that all language exhibits a degree of formulaicity (Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004), mathematics is particularly formulaic. The formulaic nature of mathematics is present not simply in the use of mathematical formulae, but also in the use of symbolic shorthand for prose. For instance, the expressions *such that*, *if and only if*, *for all*, and *there exists* each have a simplified notation ($\ni$, iff, $\forall$, and $\exists$, respectively) that removes the need to write out such expressions. This symbolic notation leads to uniformity and interpretability. The notation often comes with a specific learned meaning and verbal equivalent when read that seems to be evidence of their being processed as single units. These units may then be formulaic sequences, prefabricated language segments treated as wholes in one's memory (Wray, 1999).

A common way to identify multiword units of language of interest in a study is through frequency and range criteria. However, the specific parameters used to identify multiword units, often segments of three to five continuous words, have varied across studies. For instance, in terms of frequency, cut offs have been placed at 10 (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Simpson-Vlach & Ellis, 2010), 20 (Hyland, 2008) or 40 (Biber, Conrad, & Cortes, 2004) times per million words in lexical bundle studies, a specific subset of multiword unit investigation. In identifying range, studies have required expressions to appear across a given number of texts, such as six texts (Biber et al., 1999), a given percentage of texts, perhaps 10% (Hyland, 2008), or across a number of subcorpora, such as in four out of five genres or three out of four academic divisions (Simpson-Vlach & Ellis, 2010). The combination of such parameters identifies multiword expressions that are prevalent in the corpus, but not simply frequently used by a single author.

Studies on frequent multiword expressions in academic language abound. These have included investigations of multiword units of language across different registers such as spoken and written language (Biber et al., 2004; Simpson-Vlach & Ellis, 2010), published and student writing (Cortes, 2004), and British and American English (Liu, 2012). The resulting lists of expressions are then often categorized by structure and function to further understand the language of the corpus. Some research has even tied the use of specific multiword units to the moves and steps in research articles (e.g. Cortes, 2013). Such functional categorization is critical in not only helping us understand the language used, but also in making the research more useful for teaching (Simpson-Vlach & Ellis, 2010). Studies have shown that grammar patterns (Groom, 2005) and lexical bundles (Cortes, 2004; Hyland, 2008) vary systematically across disciplines. However, while multiword expressions may offer insight into mathematics writing, similar investigations involving the language of mathematics have been limited to classroom and educational discourse below the graduate level (e.g. Gardella & Tong, 2002; Gueunier & Larcheveque, 1972; Herbel-Eisenmann & Wagner, 2010; Herbel-Eisenmann, Wagner, & Cortes, 2010; Jingzi & Normandia, 2007; Morgan, 2005).

1.2. Phrase frames and keyness

Although continuous multiword expressions provide one phraseological path for beginning the study of mathematics research articles, another related phraseological form, the phrase frame, may provide a more comprehensive understanding of the language of these texts. The approach taken in the present study of mathematics texts is through phrase frames as Fuster-Marquez and Pennock-Speck (2015) have suggested that the identification of phrase frames distinctive to a particular genre can be useful to the teaching of English for specific purposes. Rather than being a continuous multiword expression, a phrase frame is a piece of language made up of several words and a blank. Much like simpler formulae of mathematics, phrase frames allow for an element, often a word, to be 'plugged in' a standardized functional piece of language.